

Figure 1

The novel gene as identified through RACE analysis (894 bp)

GGGAGTGGAGTGAGGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAA  
AGCTTGCCGAAGTAAAGCAAGAATGTCTTGCTCGTGGTTTGGAGACCAAGGGAATAAAG  
CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAT  
GAAGAAGATGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAAGCCCATTGAGCTCCC  
TGTCAAAGAGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAGAAAGTGG  
TGAAAATTACATCTGAAATACCACAGACTGAGAGAATGCAGAAGAGGGCTGAACGATTCA  
ATGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCT  
TCAGTTCCAACAAAAGGTCTGTCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTG  
AAGGAAAGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGAT  
GATGAGAACTGAAAAAGAGGAAGGAGCGATTTGGGATTGTCACAAGTTCAGCTGGAAC  
TGGAACACAGAGGATACAGAGGCAAAGAAGAGGAAAAGAGCAGAGCGCTTTGGGATT  
GCCTGATGAAAAGTTCCTGATACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCCTTCTTC  
TTGGTCACATATATGCCTAAATGCACAGTCATGTGCCTACGTCCTGCCTCGCAATGAGG  
GAGCATGTACCCCAGGTACATCCATGAACTGCGGCAGCAGTTTGACTTATTGCTGTTTCA  
GCTTTAAGGTTGTTGTGTTTTTGTGTTTTGATTATGTTGCTTGTTAATAAAAAAATAGAAA

A

0978476-0240

## Figure 2

### Amino acid sequence as translated from the novel gene (210 amino acids)

MATETVELHKLKLAELKQECLARGLETGKIQDLIHRLQAYLEEHAEEDVVGDETEEE  
ETKPIELPVKEEEPPEKTVDAAEKKVVKITSEIPQTERMQKRAERFNVPSLESKKAARAAR  
FGISSVPTKGLSSDNKPMVNLDKLERARFGLNVSSISRKSEDDKLLKKRKERFGIVTSSAG  
TGTTEDTEAKKRKRAERFGIA

Underlined sequences are amino acid sequences obtained by MS/MS analysis.

09788476 022404  
T07220 9488260

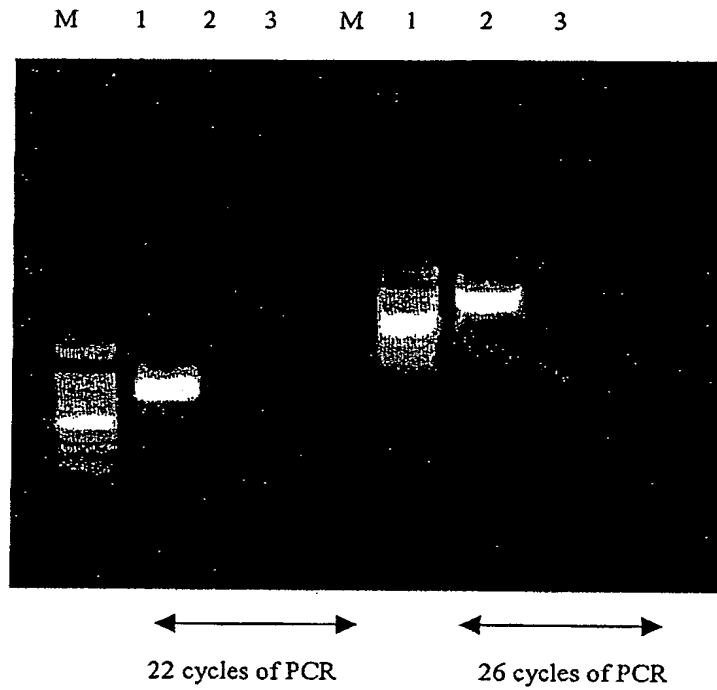
**Figure 3**

**The sequence of the novel gene amplified through long distant PCR and used to construct the expression vector (873 bp).**

TGGAGTGAGGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTT  
GCCGA ACTAAAGCAAGAATGTCTTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAAGA  
TCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAATGAAG  
AAGATGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAAGCCCATTGAGCTCCCTGTC  
AAAGAGGAAGAACCCCTGAAAAAAGTGTGATGTGGCAGCAGAGAAGAAAGTGGTGAA  
AATTACATCTGAAATACCACAGACTGAGAGAATGCAGAAGAGGGCTGAACGATTCAATGT  
ACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAG  
TTCCAACAAAAGGTCTGTCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGG  
AAAGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATG  
AGAAACTGAAAAAGAGGAAGGAGCGATTTGGGATTGTCACAAGTTCAGCTGGAAGTGA  
ACCACAGAGGATACAGAGGCAAAGAAGAGGAAAAGAGCAGAGCGCTTTGGGATTGCCT  
GATGAAAAGTTCCTGATACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCCTTCTTCTTGG  
TCACATATATGCCTAAATGCACAGTCATGTGCCTACGTCCTGCCTCGCAATGAGGGAGC  
ATGTACCCAGGTACATCCATGAACTGCGGCAGCAGTTTGACTTATTGCTGTTTCAGCTT  
TAAGGTTGTTGTGTTTTTGTGTTTGTATTATGTTGCTTGTTAAT

007846-02401  
T0220-9488260

Figure 4



09788475-0220T

Figure 5

P-151 5'-Untranslated Region

1 75  
CAGGGGCAGCAGTGATTATCTGAACTCGGATCTTTAAAATTGTGGTAGCTCTAAAGCTGATGATGTCTGGTTAGG  
\*\*\*\*\*  
76 150  
AAGTGGCTCTTGCCCGCCCCAGCCCCACCGCCAGTTCCTTAAGCCCGCCCCATGCCCTCCCAGCTTCCTCCTCA  
\*\*\*\*\*  
151 225  
TGTTTCATCGGTTTTTTCAGGGCTCCCTTCAACGCTCCCCTCTCAGTATTTAGGTCACCACTCCCTCGGCGCCCCCT  
\*\*\*\*\*  
226 300  
TTCGCTCCCACCATTTTTCTCAGCAACCCTTACAGTCTTTGCAGCTCCTACCTGCCAGCTCAGATCCCCGTCC  
\*\*\*\*\*  
301 375  
GGCTATGGGCGCGGCGCCGGCTACCACACCTGAAGTCTCCAGGAAGTAACGCCTCTCCTTCTGCCCCCTTTCCTGT  
376 450  
TGGAGGAACAGAATCAGCGCTGCCACCACCCATTGGTTGGTGGTCTGTTAATGCAGAAGCACAGTTGGTTGCCATT  
451 525  
TCTGTCTGTTTCGCAAGATACAGTGCCCGCCCCCTCTCCAGTTCCACCTTTTTGAAAGAGGTGGGGCAAGCTGCCTAG  
526 600  
AGAAGTGAGAGCGACGTCAGCTATTGACCAATGGGAAGAGCTGATGGTATGGCGTGGGAGCAAGAGTGACAACGA  
601 675  
TTGGTCAGCCTTGCATCTCTACGCCTAAGGCGGGAACCTCCTGGAGGCGGAGGCCGCGGGTGGGGGGAGTGGAGTG  
676  
AGGGGTTAACAAGATG.....P151 coding region.....

(Total length: 690 bp)

Sequence with asterisk: the 274 bp fragment

Underlined sequences are the minicistrones or uORFs before the start of the P151 coding region with the start and stop codons in bold.

P151 5'-Untranslated Region